

Evaluation of Health Service Programs and Levels of Measurement

MILTON I. ROEMER, M.D.

FOR a hundred years or more clinical medicine has applied, with varying degrees of sophistication and rigor, the method of the controlled clinical trial to test the effectiveness or value of proposed new therapies for individual patients (1). Only in the last decade or two has serious attention been given to evaluation of the various forms of organization of health services. The problems of this type of evaluation are complicated by an admixture of variables, especially involving differences between the test and control, or comparison, groups with respect to the characteristics of the persons served, the medical technology applied, or other factors outside of the form or pattern of health service organization per se.

Need for Clarification

Because of the complexities of evaluating methods of health service organization, there has been a great deal of confusion in even deciding what should be evaluated, let alone how to go about doing it (2). Many different meanings have been attributed to "evaluation," and wide disparity exists in the terminology applied to the goals of a

program, its end results, its quality, its effectiveness, its outcomes, and so on (3). The purpose of this paper is to suggest a framework for analysis of the relative values (evaluation) of various systems or subsystems of health organization (health service programs), to help clear the air and promote uniformity of terms and concepts so as to facilitate communication among investigators.

Many extensive reviews and annotated bibliographies on the problems of evaluation of health service programs have been issued in recent years. Altman and Anderson, in 1962, prepared an annotated bibliography on the evaluation of medical

Dr. Roemer is professor of health administration, University of California School of Public Health, Los Angeles. Dr. William Shonick and Dr. Carl Hopkins, University of California, Los Angeles, Dr. Jerome Schwartz, University of California, Berkeley, and Dr. Monroe Epstein, Los Angeles County Department of Hospitals, made helpful and constructive criticisms of an earlier draft of this paper. Tearsheet requests to Milton I. Roemer, M.D., University of California School of Public Health, Los Angeles, Calif. 90024.

care (4). Suchman produced a book in 1967 on evaluative research in the social sciences (5). In 1966 the Health Services Research Study Section of the Public Health Service commissioned a series of review papers on health services research. Most relevant to the issue posed here are the comprehensive accounts by Donabedian on "Evaluating the Quality of Medical Care" (6) and by Weinerman on "Research into the Organization of Medical Practice" (7). In 1967 Shapiro wrote an excellent summary on "End Result Measurements of Quality of Medical Care" (8). In August 1969 the California Center for Health Services Research issued an annotated bibliography, "Evaluating Outcomes of Health Care" (9). The World Health Organization's selected bibliography, "Methodology in Public Health Studies" (1968), also contains many references on evaluation (10). In 1969 volume 2 of "A Guide to Medical Care Administration," from the American Public Health Association's program area committee on medical care administration, appeared under the title of "Medical Care Appraisal—Quality and Utilization," prepared by Donabedian (11). A very useful compilation of readings, "Program Evaluation in the Health Fields" was also assembled in 1969 by Schulberg, Sheldon, and Barker (12).

With this wealth of reviews on the literature of health program evaluation—and there are more papers and volumes than those noted, MacTavish's bibliography (13), for example—I do not intend to add another overview, but rather offer a framework that will attempt to integrate the several approaches used into a relatively simple schema.

As organization theorists have pointed out, every system or subsystem has a set of short-term ends, which in turn become means toward more long-term ends (14). A health service program may have as its immediate end or goal the provision of certain services (for example, prenatal examinations or intensive care of patients with coronary attacks) but the long-term goal is to advance health status. There are several more links in the chain of causation, but the basic point to be recognized is that the system or program can be evaluated on various levels: short term, intermediate, and long term. Several phenomena within each level may be defined and measured.

Donabedian speaks cogently of evaluation or appraisal of the (a) structure, (b) process, and (c) outcome of a medical care program (11).

This entry to the problem is useful, especially his emphasis on importance of examining the process even though the ultimate outcome may be difficult or impossible to measure. My attempt, in a sense, is to refine this typology somewhat further, in order to build a framework into which the whole spectrum of evaluation methodologies may logically be fitted.

If the focus is on evaluation of health service programs (that is, mechanisms of organization of health services in their many aspects) it may be helpful to think of all the consequences as a chain of effects at different levels of depth. Regardless of the immediate short-term ends or goals of a health program, it must ultimately be judged or evaluated by its success in saving lives or reducing disability of advancing health status in some way. Only when the attainment of that ultimate goal becomes difficult to measure or to attribute to a specific programmatic cause, which is frequently true, must we take recourse to evaluation based on less ultimate effects (15).

Levels of Evaluation

Health status outcomes. Ideally, health planners would like to know the effect of any pattern of organization of health services, whether old, new, or projected, in terms of health status changes in the target population. Many studies have used this type of outcome measure with varying degrees of sophistication in ruling out secondary variables.

On the crudest level of a total population, for example, one may compare a large system like the British National Health Service with the U.S. health scene. Observing a higher life expectancy in Great Britain than in the United States, one might conclude that the net outcome of the British National Health Service is superior to that of the pluralistic U.S. system, or "nonsystem" (16). But such a conclusion would be unwarranted without considering the effects of diverse living conditions, genetic factors, and scores of epidemiologic variables that can influence death rates and life expectancy in the two nations, quite aside from their health service systems. Nevertheless, even this crude comparison provides a clue for more searching types of measurement of the effects of the two systems of health service at the deepest level of evaluation; namely, the outcome in health status.

More sophisticated evaluative studies are illustrated by comparisons of the membership of the

Health Insurance Plan of Greater New York with the rest of the New York population, matched for sociodemographic characteristics. In the early 1950's an important study showed lower perinatal mortality in a population eligible for this prepaid group practice program (17), and in the 1960's a study showed a lower death rate among indigent aged (old-age assistance recipients) enrolled in the plan (18), compared in both instances with matched populations entitled to traditional medical care. The elaborate tasks of sampling, randomization, data analysis, and so on in studies of this type need not be reviewed here.

Health status outcomes have also been applied in comparative studies of populations actually served in varying medical settings, most frequently in hospitals of different types. Lipworth and co-workers found lower disease-specific case fatality rates in British teaching hospitals compared with non-teaching ones (19). John Thompson and colleagues compared perinatal mortality as an indicator of obstetrical care in two U.S. Air Force hospitals (20). I found lower postoperative deaths for certain surgical procedures in large, compared with small, hospitals in Saskatchewan (21).

In all such hospital-based studies, one must make adjustments for the varying severity of cases, and hence risk of death, in different hospitals; and in a 1968 paper two colleagues and I offered a statistical approach to the solution of this problem (22). If such a statistical adjustment can be perfected, we will have a much firmer basis for judging a hospital's overall effectiveness than the brief inspections of input (hospital resources, policies, practices) on which the Joint Commission on Accreditation of Hospitals or the State hospital licensure authorities now depend.

Beyond mortality data are many other measures of the ultimate outcome of a health service program, in terms of health status, that may be applied, either to total populations eligible or to persons definitely served by the program. Life expectancies, based on modified life table techniques, have been applied to measure the effect of a county public health program. This method corrects for the problem of higher morbidity rates among older persons with chronic disease who are kept alive by active medical care (23). More sensitive than this are various measures of recovery from illness or days of disability, such as absenteeism from work or school, restricted activity days, or days in bed, of persons eligible for one program as compared with another (24).

Formulated more positively, health status outcomes may be reflected in measurements of the capacities of persons to function, as applied by Sidney Katz and his colleagues in studies of rehabilitation of the aged sick (25,26). The effectiveness of family planning programs may be evaluated in terms of subsequent birth rates. There are scores of specific measurements of recovery from certain diseases, improved physical or mental functioning, and other phenomena that may be and have been applied in outcome evaluation of specialized programs (27). Sanazaro and Williamson have delineated a set of six patient end-results for judging the outcome of cases reaching the attention of specialists in internal medicine (28).

This type of health status outcome is the usual end point in clinical trials of new drug therapies or new preventive services, like the Salk anti-polio myelitis vaccine or the fluoridation of water supplies to reduce the rate of dental caries. Its application to evaluation of health service programs, however, is complicated by so many variables in the characteristics of the populations eligible or served, the diseases involved, and numerous environmental divergencies that, in practice, it is difficult and costly to apply. To adjust properly for all these confounding influences requires very large or highly selected samples, long periods of observation, and elaborate methodologies. As a practical matter, therefore, evaluation of health service programs must often resort to measurements of effect that are less ultimate in the chain of influences. Next to health status would be the level that may be described as the estimated quality of service.

Estimated quality of service. This level of program evaluation is a component of Donabedian's "process." It can be applied, by definition, only to examination of services actually rendered rather than to the experience of a total population, some members of which may receive no services at all. The measurement rests on the assumption that at any time and place there is a scientific consensus among widely acknowledged experts on what constitutes good or high-quality health service. The consensus typically, though not always, rests on a body of empirical data. The task then is to call upon an expert observer to examine, directly or indirectly, the services actually provided in a program and make judgments on the degree to which the services coincide with these accepted standards of merit. The judgment may be scaled from high to low, may be given a

numerical score, or may be subdivided along different dimensions of service.

The most common application of this level of evaluation has been to hospital services through study of patient records. Known generally as the medical audit technique, it has been applied extensively by Lembcke (29) and Rosenfeld (30) as well as others with methodological variations that need not be reviewed here. Numerous investigations have been made of such outcomes as rates of appendectomy (for a physician or for a hospital) associated with nonpathological findings or the proportions of post mortem findings that did not confirm the original diagnosis. The widely publicized study of the quality of medical care received by members of the Teamsters Union under a health insurance plan was based on this audit technique (31). With somewhat greater difficulty, the study of written records as a basis for evaluation has been applied also to services for ambulatory patients (32).

Because of the many possible inadequacies of the written record as a reflection of what was actually done (that is, errors of commission or omission in the record), the quality of services may be judged also by visual observation. This technique was used in the well-known studies of general medical practice by Peterson and his colleagues in North Carolina (33) and by Clute in Canada (34). Visual observations of patients' mouths have likewise been applied in evaluative studies of dental service programs. Prescribing practices of physicians have also been examined as indicators of their quality of performance (35).

There are endless ramifications to the types of judgmental observations that may be made at this level of evaluation. Instead of applying a standard of excellence, the performance in a particular program may be compared with an average of many such programs, as is the strategy of the Commission on Professional and Hospital Activities (36). The old appraisal schedule of local public health programs used by the American Public Health Association was applied largely in this way (37).

Apart from the fallibilities in judgment of any "expert," this whole level of measuring results is often difficult to apply because of the inaccessibility of records or other objects of observation, because of the expense involved, or for other reasons. Therefore, program evaluation of results must often take recourse to a third level of meas-

urement: the quantity of services provided. This is another facet of Donabedian's concept of process.

Quantity of services provided. The basic assumption of this evaluative level is that certain types of health service (not all types) may be regarded as generally beneficial for people, so that a higher rate of providing these services to a population is deemed more favorable than a lower rate. One can immediately think of exceptions to this generalization, but the argument for its usual validity rests on the entire literature and knowledge of the field of scientific clinical medicine. In general, other things being equal, it is assumed that a health service program which yields a high rate of contact between patients and physicians is better than one which yields a lower rate. (This view has been widely held since about 1912; Dr. Reginald Fitz of Boston set that year as the date after which an encounter between a patient and a doctor yielded a better than 50-50 chance of benefit for the patient.) The whole extension of health insurance programs, for example, has been advocated on the basis of the statistical demonstration that insured persons (of given age, sex, and socioeconomic status) get more health services per year than noninsured persons with the same characteristics (38).

The quantity of health services provided to a population by a program, or the utilization rates if viewed from the standpoint of the recipients, may be of many different categories and subcategories. Most elementary is the determination of the percentage of a stated population reached (that is, provided one or more units of service by the program) during a year. For physician contacts this is often between 50 and 75 percent, even when costs are covered by an insurance or public program (39).

Beyond this, one may determine the rate of receipt by the eligible population of ambulatory medical services, hospitalizations, prescribed drugs, dental services, and so on. Within ambulatory medical services, one may measure preventively oriented services, like physical examinations or immunizations, or many types of diagnostic or treatment procedures. Dental services may include rates of prophylaxes, fillings, extractions, prostheses, and others. Hospitalization may be measured by cases or admissions, by days of care, by diagnostic category, and so on. All these measurements, of course, depend on minimally adequate medical records and a clear definition of terms (40).

Such data on the quantity of various types of service received by an eligible population are clearly a program consequence although their appraisal as good, fair, or poor requires further interpretation. When rates of service provided are markedly different from certain well-known experiences, the judgment is easier; for example, we know that the U.S. population as a whole utilizes hospital services at the rate of about 1,100 days per 1,000 persons per year and sees physicians at the rate of about five contacts per person per year. If we then observe that the annual utilization rates in, let us say, the State of West Bengal, India, are 200 days per 1,000 persons and one physician contact per person, we need not hesitate to conclude that the West Bengal health service system has serious deficiencies. When the differentials are slight, however, we cannot usually make value judgments; yet we can draw simple conclusions on the quantitative effects of a program that can be useful for planning purposes.

When certain medical procedures have been clinically demonstrated to be of dubious benefit (for example, tonsillectomies or uterine suspensions), a high rate of their performance can reflect low programmatic quality. On the other hand, certain procedures may be deemed of generally high value, such as immunizations for the young or proctoscopic examinations of aged persons, and these rates therefore have other meaning. Rates of hospitalization as a whole under different types of insurance plans have been extensively studied as an outcome reflecting possible abuse or overuse of expensive facilities, as well as the compensatory value of out-of-hospital services to the ambulatory patient (41).

Focusing only on the persons actually receiving services within a program, one may undertake other measurements useful for evaluation. The time spent per patient, or the number of patients seen per physician-hour, is a useful measurement. Within a physician's practice, the proportion of patients given injections or subjected to certain diagnostic tests may be an evaluative index. In a dental program, the ratio of fillings to extractions is widely regarded as reflecting a preventive orientation. The ratio of prescribed to nonprescribed drugs consumed by a population is another index reflecting quality. The rate of noncompliance with medical orders or advice is a special form of program measurement that also has obvious qualitative implications. While interpretations of the meaning of these various quantitative rates, for

specified types of health service, must obviously be made with caution, such measurements constitute a level of evaluation that permits interprogram comparisons (42).

Such relatively simple counts as these may not even be possible, however, if proper records are not kept in a health service program. A common difficulty is the lack of knowledge of the size of the eligible population so that, with no clear denominator, basic rates cannot be calculated at all, and only proportions of patients getting certain services can be measured. Evaluation of a health program may be most feasible, therefore, at still another level: the attitudes of the persons whom the program is intended to serve.

Attitudes of recipients. Combining quantity and quality, in a sense, is the measurement of a health service outcome that is based on the attitudes of persons entitled to or actually receiving the service. Without knowing the quantitative rates of services provided or their estimated quality (as judged by professional experts), one can ask people how they feel about the program. Many evaluative studies have been based on this type of survey measurement. The impact of diverse types of health insurance plans on persons enrolled has been studied in this way among State government employees in California (43), among insured persons in New York City (44), and in other settings. Opinion surveys of the attitudes of British people before and after the National Health Service have been used to evaluate that large program (45).

Although the judgment of a program member or patient may often be superficial and faulty, this method assumes that such judgment has some validity and will certainly reflect gross problems in a program. For the humanistic and personal aspects of health service, this level of evaluative measure is probably more cogent than any other. Moreover, this type of measurement is probably the best approach to quantification of such frequently espoused criteria for good medical care as accessibility, acceptability, continuity, comprehensiveness, sensitivity, and the like.

Within the population actually receiving services, a quantification of grievances may also be a tool of evaluation. Hospitals and health insurance plans often invite patients to comment in writing on the services they have received, using the rate of specific complaints as a key to program improvement. In more extreme form, a study of the rate of malpractice suits in a series of California

hospitals found this measure to reflect the degree of rigor in the organization of the medical staff (46).

As democratic concepts become more embodied in the provision of health service and as the sophistication of people about medical science broadens, this level of evaluation can become increasingly important. Witness the ferment in the nationwide Medicaid program, associated not only with rising costs but also with documented complaints of poor people about the nature of the services they get (47). This type of measurement need not require medical records nor the other elaborate forms of data necessary for the three previous levels, but it usually requires population surveys through interviews, questionnaires, or other means, which must be done with care and may be quite costly. On a level still further from the goal of improvement in health status, therefore, one may draw inferences about the operation of a program by measuring the various attributes of the personnel and physical resources made available in it. These are equivalent to Donabedian's concept of structure.

Resources made available. While human and physical resources are ordinarily thought of as inputs rather than outputs of a system, these resources require much effort for production and distribution, so that they may also be viewed as consequences. This concept may be seen clearly in an underdeveloped country, where the results of a national program of rural health improvement may be measured by the simple ratio to population of physicians, nurses, hospital beds, and so on, achieved in rural areas (48). In the United States an immediate result of the national Hill-Burton program is the number or ratio of hospital beds established in regions formerly undersupplied, and similar measures may be used for personnel trained and working as an achievement of various health manpower development programs (49). These ratios, it may be noted, apply theoretically to total populations rather than to patients reached.

The assumption, of course, is that personnel and physical resources result in services, just as the services, in turn, are presumed generally to yield benefits for health. Anyone can spot the possible fallacies in these assumptions, and yet they are more likely to be valid than not. For years the local public health promotional program of the Public Health Service reported its progress in terms of the number of counties (among the

3,070 in the nation) served each year by full-time health departments (50). The assumption was that these structural resources led to certain services which, in turn, reduced communicable diseases, infant mortality, and so on. There was enough independent scientific evidence of the benefits of immunizations, sterilization of baby formulas, early detection of tuberculosis, and so on, to justify these assumptions in a broad sense. More ultimate measures of the results of local health department programs are sought, naturally, but even at this fifth level certain probabilistic conclusions are warranted. In some situations, no more satisfactory evaluative data may be obtainable.

Within this evaluative level of resources made available, measurements may be further refined along qualitative lines. One may count the kinds of physicians available, for example, distinguishing general practitioners and qualified specialists (51). One may define the range of equipment and the scope of services offered in hospitals and health centers. The whole literature of clinical medicine justifies the assumption, if other data are lacking, that a fully trained surgeon is likely to achieve better surgical results for his patient than a general practitioner. While exceptions may occur, in a complex medical situation it is probable that a professional nurse will be more helpful to a patient than a vocational nurse. Broadly speaking, a health program may be subjected to an administrative audit in which all its resources are defined and their manner of functioning is described; certain operating procedures imply superior or inferior service.

Beyond this fifth level of evaluation of health service programs, there is still another type of question to be asked: What are the costs of the program? All five levels discussed are measures of benefits but they tell us nothing about the costs and hence of the cost-benefit ratios. Since health service resources are always limited, it is reasonable to attempt to achieve a stated outcome at the lowest possible cost, and the measurement of this cost may therefore be regarded as another type of evaluation.

Costs of the Program

If a stated health objective at any of the five levels can be reached by one method at a lower cost than by another method, there is greater efficiency and higher value in the first method. It means that more money or resources would then be left for meeting other needs or demands (52).

The costs of a program, while a type of evaluation, are along a dimension different from that of the five levels of benefits discussed. For any quantity of benefits at any of the five levels, there may be a range of costs. The difficulty is to be certain that comparative cost measurements are being applied to health service programs that do, indeed, reach the same results. If not, there must at least be some uniform units of measurement of results, such as days of disability incurred or number of dental services provided, under alternative systems, so that cost-benefit analyses and comparisons can be made (53). How much should be spent for gaining a stated objective is a matter for social policy decision, and choices must always be made among large sectors like health, military affairs, education, housing, and so on. Within the health sector alone, the choices are difficult enough, but between these large sectors the cost-benefit calculations are so formidable that they are seldom even attempted, and the decisions are usually left to political judgment.

Within the health services, costs can be calculated in several ways. As at the other levels of evaluation, the cost measurement may be applied to the total population eligible or to the population actually served. The first dimension requires calculation of the cost over time per person eligible for service; the second dimension requires only determination of the cost of services actually rendered, such as a physician visit or a hospital day. By either type of measure, comparisons of cost may be made between different methods of seeking to achieve the same goal; for example, water fluoridation versus periodic topical fluoride applications to children's teeth or group medical practice with salaried physicians in contrast to solo practice with fee-for-service remuneration (54).

These fiscal measurements are complex because hidden costs must not be overlooked. If the cost of an organized home care program is to be compared with equivalent long-term hospital care, one must not ignore the expenses incurred by a family in keeping a sick member at home (55). One must also not overlook administrative costs in a program's operation; for example, a high rate of personnel turnover in a clinic creates hidden costs for training new employees or reduced efficiency until new personnel learn their tasks. In hospital cost calculations, the shares of professional education and research costs that are properly accountable to patient care are perennially debated. If the laboratory in a hospital is understaffed, a bottle-

neck may be caused in the flow of patient care, leading to longer durations of stay; this administrative problem might not be reflected in per diem costs but only in cost per hospital case (56).

In any event, cost figures are a far cry from health status as measurements of the ultimate outcome of a health service program, but they are nevertheless relevant to many larger questions of social policy.

Comment

In this review of five levels of benefit evaluation for health service programs, and a sixth level of cost evaluation, each level is presented along a gradient of depth or ultimacy. In the logic of the means-and-ends chain, this is believed to be generally valid, but as a practical matter there are circumstances in which a less ultimate level of evaluation may actually be more desirable than a more ultimate one. Thus, for example, the secondary variables, outside of the health service program, influencing death rates (level 1) may be so numerous and so difficult to adjust for, that the estimated quality (level 2) or even the simple quantity (level 3) of services provided by the program may be more reliable measures of its effects. For another example, the medical records of services in a program may be so inadequate that an interview with recipients on their attitudes (level 4) toward it, despite all the fallacies of the layman's judgment, may be more reliable than estimates of quality by expert review of written charts.

Because of these difficulties, many efforts to evaluate health service programs properly seek to measure two or more different levels of results at the same time. The studies of old-age assistance clients served by the Health Insurance Plan of Greater New York, compared with the traditional patterns (18), obtained data on the quantity of different services provided as well as on health status (mortality rates). Our research on diverse health insurance plans in California is determining the quantity of services provided, the attitudes of recipients, and the costs or expenditures under each plan, although we are not getting health status measurements (57).

The most useful strategy for evaluation would be to determine the effects of health service programs at all five levels plus their costs. Until that can be achieved, we should realize that measurements at any one or more of the six evaluation levels contribute something to our understanding

Level of evaluation	Eligible population	Patients served
1 Health status		
2 Estimated quality		
3 Quantity of service		
4 Attitudes		
5 Resources		
6 Costs		

of health service systems and are far more useful than evaluations based on intuition or speculation.

There are endless methodological problems in sampling, data collection, analysis, and so on that I have not discussed in this paper. It should be pointed out, however, that the health service program to be studied may have varying degrees of complexity, and the problems of evaluation are corresponding. Evaluation may be attempted of an entire national health service system or, for example, of a particular prenatal clinic, or of a large series of program complexities in between. Intermediate complexities might be illustrated by a regional network of hospitals, a health insurance plan, or an air pollution control program.

The more complex the program examined, the more numerous generally are the secondary variables that must be adjusted for. The microsystem questions can usually be answered more rapidly and less expensively than the macrosystem questions. Ease of solution, however, seldom corresponds to the social importance of a question, and we should avoid tackling certain evaluative problems just because they are easily soluble, unless at the same time they are socially salient.

In summary, then, one can conceptualize a matrix of evaluation of health service programs. Along one axis would be the six levels of results that may be measured, and along the other axis

the applicability of the measurement to a total eligible population or to only the patients actually served. Within each dimension there would be programs of varying degrees of complexity that may be summarized as macro or micro systems. This matrix could be schematized as shown at left.

Within each of these conceptual cells, evaluation requires the comparison of measurements of at least two entities, defined either across time (before and after) or across space (the model of test and control groups). Studies within any of these conceptual cells can be useful for program evaluation, which can facilitate a rational planning of health services.

REFERENCES

- (1) Shryock, R. H.: American medical research past and present. Commonwealth Fund, New York, 1947.
- (2) B'tesch, S.: International research in the organization of medical care. *Med Care* 4: 41-46, January-March 1966.
- (3) Kerr, M., and Trantow, D. J.: Defining, measuring, and assessing the quality of health services. Perspectives and a suggested framework. *Public Health Rep* 84: 415-424, May 1969.
- (4) Altman, I., and Anderson, A. J.: Methodology in evaluating the quality of medical care: An annotated selected bibliography, 1955-61. University of Pittsburgh Press, Pittsburgh, Pa., 1962.
- (5) Suchman, E.: Evaluative research: Principles and practice in public service and social action programs. Russell Sage Foundation, New York, 1967.
- (6) Donabedian, A.: Evaluating the quality of medical care. *Milbank Mem Fund Q* 44: 166-206, July 1966.
- (7) Weinerman, E. R.: Research into the organization of medical practice. *Milbank Mem Fund Q* 44: 104-145, October 1966.
- (8) Shapiro, S.: End result measurements of quality of medical care. *Milbank Mem Fund Q* 45: 7-30, April 1967.
- (9) Klein, B.: Evaluating outcomes of health care: An annotated bibliography. California Center for Health Services Research, Los Angeles, August 1969.
- (10) World Health Organization: Methodology in public health studies: A selected bibliography. OMC/RES/68.5. Geneva, December 1968.
- (11) Donabedian, A.: Medical care appraisal—quality and utilization: A guide to medical care administration. Vol. 2. American Public Health Association, New York, 1969.
- (12) Schulberg, H. C., Sheldon, A., and Baker, F., editors: Program evaluation in the health fields. Behavioral Publications, New York, 1969.
- (13) MacTavish, C. F.: Assessment of the quality of

- medical care: An annotated bibliography. American Rehabilitation Foundation, Minneapolis, Minn., February 1968.
- (14) Simon, H. A.: *Administrative behavior*. Macmillan Co., New York, 1961, pp. 62-66.
 - (15) Hilleboe, H. E., and Schaefer, M.: Evaluation in community health: Relating results to goals. *Bull NY Acad Med* 44: 140-158, February 1968.
 - (16) World Health Organization: *Vital statistics and causes of death: World health statistics annual*. Vol. 1. Geneva, 1968.
 - (17) Shapiro, S., Jacobziner, H., Densen, P., and Weiner, L.: Further observations on prematurity and perinatal mortality in a general population and in the population of a prepaid group practice medical care plan. *Am J Public Health* 50: 1304-1317, September 1960.
 - (18) Shapiro, S., et al.: Patterns of medical use by the indigent aged under two systems of medical care. *Am J Public Health* 57: 784-790, May 1967.
 - (19) Lipworth, L., Lee, H., and Morris, J. N.: Case fatality in teaching and non-teaching hospitals, 1956-1959. *Med Care* 1: 71-76, April 1963.
 - (20) Thompson, J. D., Marquis, D. B., Woodward, R. L., and Yeomans, R. C.: End-result measurements of the quality of obstetrical care in two U.S. Air Force hospitals. *Med Care* 6: 131-143, March-April 1968.
 - (21) Roemer, M. I.: Is surgery safer in larger hospitals? *Hosp Manage* 87: 35-37, 50, 77, 101, (1959).
 - (22) Roemer, M. I., Moustafa, A. T., and Hopkins, C. E.: A proposed hospital quality index: Hospital death rates adjusted for case severity. *Health Serv Res* 3: 96-118, Summer 1968.
 - (23) Sanders, B.: Measuring community health levels. *Am J Public Health* 54: 1063-1070, July 1964.
 - (24) Sullivan, D.: Conceptual problems in developing an index of health: Vital and health statistics. PHS Publication No. 1000, Ser. 2, No. 17. U.S. Government Printing Office, Washington, D.C., May 1966.
 - (25) Katz, S., et al.: Studies of illness in the aged: The index of ADL, a standardized measure of biological and psychosocial function. *JAMA* 185: 914-919, Sept. 21, 1963.
 - (26) Katz, S.: Practical experience in evaluating a health service program. *In* Synopsis of proceedings, outcomes meeting I, California Center for Health Services Research, Los Angeles, May 1969.
 - (27) Hagner, S., LoCicero, V., and Steiger, W.: Patient outcomes in a comprehensive medicine clinic. *Med Care* 6: 144-156, March-April 1968.
 - (28) Sanazaro, P. J., and Williamson, J. W.: End results of patient care: A provisional classification based on reports by internists. *Med Care* 6: 123-130, March-April 1968.
 - (29) Lembcke, P. A.: Evaluation of the medical audit. *JAMA* 199: 543-550, Feb. 20, 1967.
 - (30) Rosenfeld, L. S.: Quality of medical care in hospitals. *Am J Public Health* 47: 856-865, July 1957.
 - (31) Ehrlich, J., Morehead, M. A., and Trussell, R. E.: The quantity, quality and costs of medical and hospital care secured by a sample of teamsters families in the New York area. Columbia University, New York, 1962.
 - (32) Kroeger, H. H., et al.: The office practice of internists: The feasibility of evaluating quality of care. *JAMA* 193: 371-376, Aug. 2, 1965.
 - (33) Peterson, O. L., et al.: An analytical study of North Carolina general practice. *J Med Educ* 31: 1-165, December 1965.
 - (34) Clute, K. F.: The general practitioner: A study of medical education and practice in Ontario and Nova Scotia. University of Toronto Press, Toronto, 1963.
 - (35) Furstenberg, F. F., et al.: Prescribing as an index to quality of medical care. *Am J Public Health* 43: 1299-1309, October 1953.
 - (36) Slee, V. N.: Uniform methods of measuring utilization. *In* Utilization review: A handbook for the medical staff. American Medical Association, Chicago, 1965.
 - (37) Hiscock, I. V.: *Community health organization*. Ed. 4. Commonwealth Fund, New York, 1950, p. 242.
 - (38) Anderson, O. W., and Feldman, J. J.: Family medical costs and voluntary health insurance: A nationwide survey. McGraw-Hill Book Co., New York, 1956.
 - (39) Darsky, B. J., Sinai, N., and Axelrod, S. J.: Comprehensive medical services under voluntary health insurance: A study of Windsor medical services. Harvard University Press, Cambridge, Mass., 1958, p. 62.
 - (40) Densen, P.: Some practical and conceptual problems in appraising the outcome of health care services and programs. Paper presented at Health Services Outcomes Conference, Los Angeles, Dec. 1, 1969.
 - (41) Klarman, H. E.: Controlling hospital use through organization of medical services. *In* Where is hospital use headed? Paper presented at the Fifth Annual Symposium on Hospital Affairs, University of Chicago, 1962, pp. 55-63.
 - (42) Roemer, M. I.: Research in administrative medicine: Comparative analysis of systems of health service organization. *In* Research in social welfare administration. National Association of Social Workers, New York, 1962, pp. 72-81.
 - (43) Watts, M., et al.: A special report of the medical and hospital advisory committee to the board of administration. State Employees Retirement System, Sacramento, Calif., June 1964.
 - (44) Anderson, O. W., and Sheatley, P. B.: Comprehensive medical insurance: A study of costs, use, and attitudes under two plans. Health Information Foundation, Chicago, 1959.
 - (45) Gemmill, P. F.: An American report on the National Health Service. *Br Med J (supp.)* 5: 17-21, July 1958.
 - (46) Blum, R. H.: Hospitals and patient dissatisfaction. Technical Report. California Medical Association, San Francisco, 1958.
 - (47) U.S. Department of Health, Education, and Welfare: Recommendations of the task force on

- Medicaid and related programs. U.S. Government Printing Office, Washington, D.C., November 1969.
- (48) Roemer, M. I.: The rural health services scheme in Malaysia. World Health Organization, Western Pacific Regional Office, Manila, February 1969.
- (49) National Commission on Community Health Services: Health manpower: Action to meet community needs. Public Affairs Press, Washington, D.C., 1967.
- (50) U.S. Public Health Service: Directory of local health units, 1966. PHS Publication No. 118. U.S. Government Printing Office, Washington, D.C., 1966.
- (51) Mott, F. D., and Roemer, M. I.: Rural health and medical care. McGraw Hill Book Co., New York, 1948.
- (52) McCaffree, K.: The cost of mental health care under changing treatment methods. *Am J Public Health* 56: 1013-1025, July 1966.
- (53) Klarman, H.: Present status of cost-benefit analysis in the health field. *Am J Public Health* 57: 1948-1954, November 1967.
- (54) Williams, J. J., et al.: Family medical care under three types of health insurance. Foundation on Employee Health, Medical Care, and Welfare, New York, 1962.
- (55) Rogatz, P., et al.: Organized home medical care in New York City. Harvard University Press, Cambridge, Mass., 1956.
- (56) Lave, J.: A review of the methods used to study hospital costs. *Inquiry* 3: 57-81 (1966).
- (57) Sasuly, R., and Roemer, M. I.: Health insurance plans: A conceptualization from the California scene. *J Health Hum Behav* 7: 36-44, spring 1966.